CHEMICAL ENGINEERING TRANSACTIONS
VOL. 51, 2016
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.aidic.it/cet
Guest Editors: Tichun Wang, Hongyang Zhang, Lei Tian

ISBN 978-88-95608-43-3; **ISSN** 2283-9216
DOI: 10.3303/CET1651098

Research on Intrusion Detection System Based on Improved PSO-SVM algorithm

Bin Tan[*a], Yang Tan[b], Yuanxu Li[c]

[a] College of Computer Science , Jinjiang College, Sichuan University
[b] College of Computer Science and Engineering, Chongqing University of Science and Technology
[c] College of Foreign Languages, Jinjiang College, Sichuan University.
tanb_ice@163.com

With the rapid development of Internet, the network topology structure becomes more and more complex, so that the monitoring of network attack has become quite difficult. The traditional passive defence strategy has been unable to meet the demand of network information security. How to effectively detect and prevent the network intrusion have become an important matter in the field of computer security. The efficient intrusion detection system can reduce the false positive rate of the system, and improve the classification accuracy. This paper firstly introduces the intrusion detection system and detection data set. On this basis, this paper proposes an intrusion detection method based on improved PSO-SVM. The support vector machine can ensure that classifier has high classification accuracies. Secondly, we use PSO method to determine the important parameters of the SVM algorithm, such as the RBF kernel parameter, penalty parameter and insensitive loss error. Then, the improved PSO method can find the optimal value of the SVM. At this time, the error sum of squares of the SVM model has a minimum value, and the model has a fast convergence speed. Finally, because the training data sets of DoS and Probe are accounted for a larger proportion of all attacks, we use the IPSO-SVM classification algorithm for them, and have a test to the intrusion detection. The experimental results show that the overall performance of the proposed detection algorithm is very high. It has a strong ability to identify the characteristics of intrusion, and can provide the intrusion detection services for virtual environment.

1. Introduction

With the rapid development of Internet, the network topology structure becomes more and more complex, so that the monitoring of network attack has become quite difficult. The traditional passive defence strategy has been unable to meet the demand of network information security. How to effectively detect and prevent the network intrusion have become an important matter in the field of computer security. Therefore, to develop an efficient intrusion detection system is imminent. At present, many effective intrusion detection classification models are proposed. Yang et al., (2010) propose a new model based on network protocol analysis and decision tree mining technology. According to analyze the protocol type of data packet, this model can determine the optimal DT algorithm for intrusion detection test. In addition, this method has good detection effect in high-speed network environment. There are also other learning algorithms used in the field of intrusion detection, such as support vector machine(Xu and Li, 2012; Li et al., 2012), genetic algorithm(Wang et al., 2010; Li et al., 2003) and artificial neural network (Liu, 2014; Zhang, 2012). Dorothy E(1987) describes a real-time intrusion detection expert system that can detect intrusion, penetration and other illegal operations. Martin Roesch(1999) describes the intrusion detection system, represented by Snort, which is based on rule matching. Lee W (1999) presents the data mining framework which is used to construct the adaptive intrusion detection model. The traditional intrusion detection systems often need to update the intrusion detection system, so as to cope with the new attacks. However, many intrusion detection systems are relying on the expert experiences, and any changes are costly and slow. In order to solve the problem of intrusion detection system, Koc et al., (2012) present the hidden naive Bayesian model. Shon et al., (2006) present a method

Please cite this article as: Tan B., Tan Y., Li Y.X., 2016, Research on intrusion detection system based on improved pso-svm algorithm, Chemical Engineering Transactions, 51, 583-588 DOI:10.3303/CET1651098

which is used the genetic algorithm to select the feature of TCP/IP data packets. Zhang et al., (2008) present a new random forest method, and it is applied in the intrusion detection system.
To sum up, this paper firstly introduces the intrusion detection system and detection data set. On this basis, this paper proposed an intrusion detection method based on improved PSO-SVM. The support vector machine can ensure that classifier has high classification accuracies. Secondly, we use PSO method to determine the important parameters of the SVM algorithm, such as the RBF kernel parameter, penalty parameter and insensitive loss error. Then, the improved PSO method can find the optimal value of the SVM. At this time, the error sum of squares of the SVM model has a minimum value, and the model has a fast convergence speed. Finally, because the training data sets of DoS and Probe are accounted for a larger proportion of all attacks, we use the IPSO-SVM classification algorithm for them, and have a test to the intrusion detection.

2. Intrusion detection system

With the rapid development of Internet, the network topology structure becomes more and more complex, so that the monitoring of network attack has become quite difficult. The traditional passive defence strategy has been unable to meet the demand of network information security. The reasons are as follows:
(1) The intruder can bypass the firewall to find the open back door.
(2) The intruder may be within the firewall.
(3) Due to the performance limit, the firewall usually cannot provide the real-time intrusion detection.
Therefore, it requires the intrusion detection system as a supplement. The system can timely find various attempts and behaviours of attack, and respond to them. In this way, the intrusion detection system and many security products constitute a three-dimensional security system. The intrusion detection system is one of the most important parts in security system. Figure 1 is the framework of security system,

3. The classification process of PSO-SVM

3.1 SVM model

First of all, the algorithm should be given training sample set which is $T=\{x_i, y_i\}$, $(i=1,2,\cdots,n)$. Where, n is the number of training samples, i is the training sample, x_i is the input vector, and the corresponding expected value is y_i. The basic idea of SVM regression is to use the nonlinear function $\phi(x)$ to map the data of input space to high dimensional feature space. Then, we can use the linear regression which is the $f(x)=\omega\phi(x)+b$ in this space. Where, ω is the weight vector and b is the deviation. The parameters ω and b can be obtained by minimizing a regularized risk function $R(C)$:

$$R(C)=\frac{C}{n}\sum_{i=1}^{n}L_{\varepsilon}(y)+\frac{1}{2}\|\omega\|^2 \tag{1}$$

$$L_{\varepsilon}(y)=\begin{cases}|f(x)-y|-\varepsilon & |f(x)-y|\geq\varepsilon\\ 0 & |f(x)-y|<\varepsilon\end{cases} \tag{2}$$

Where, $\|\omega\|^2$ is the structure risk, $L_{\varepsilon}(y)$ is insensitive loss function, ε is insensitive loss error, and its value directly affects the number of support vector. C is the penalty parameter, and it controls the degree of punishment. At this point, we introduce non negative slack variables which are ξ and ξ^*. They are used to measure the deviation degree of sample. Then, the optimization problem of formula (1) can be expressed by the following formula:

$$\min[\frac{1}{2}\|\omega\|^2+C\sum_{i=1}^{n}(\xi+\xi^*)]$$

$$\text{s.t.}\begin{cases}y_i-(\omega\phi(x)+b)\leq\varepsilon+\xi_i & \xi_i\geq 0\\ y_i-(\omega\phi(x)+b)\geq-\varepsilon-\xi_i^* & \xi_i^*\geq 0\end{cases} \tag{3}$$

The kernel function can use the function of the input space as the inner product operation in high dimensional space, so as to avoid the dimension disaster. The kernel function has many kinds, such as radial basis kernel

function, Gauss kernel function, polynomial kernel function. This paper chooses the RBF kernel function, and it can be described as below:

$$K(x,x_i)=e^{-\frac{\|x-x_i\|^2}{2\sigma^2}} \tag{4}$$

Where, σ is the radial basis kernel parameter. Then, we use the RBF kernel parameter σ and penalty parameter C and insensitive loss function ε as the training parameters of SVM.
SVM regression function is:

$$f(x)=\sum_{i=1}^{n}(a_i-a_i^*)K(x,x_i)+b \tag{5}$$

3.2 PSO model

Particle swarm optimization algorithm is a swarm intelligence optimization algorithm which is mainly used to search for the global optimal solution. PSO is initialized to a group of random particles, and the position of the particle is an potential solution of optimization problem in the search space. The velocity of particle determines the direction and distance of the flight, and all particles have a fitness function. All the particles follow the current optimum particles in the solution space, and find the optimal solution by the iterative algorithm. At each iteration, the particle is updated by tracking individual extremum $pbest$ and global extremum $gbest$.
Next, we search for a group with u particles in the Q-dimensional space. Set the position of the particle g ith is $D=(d_{g1},d_{g2},\cdots,d_{gn})$, and the speed is $V=(V_{g1},V_{g2},\cdots,V_{gn})$. The formula (6) and (7) can be used to calculate the position and the speed.

$$v(k+1)=\omega v(k)+c_1 rand_1(pbest(k)-h(k))+c_2 rand_2(gbest(k)-h(k)) \tag{6}$$

$$h(k+1)=h(h)+\beta v(k+1) \tag{7}$$

Where k is number of evolution, v is the velocity of particle, and h is the position of the particle. The ω is the inertia weight which is used to balance the global search and local search. The β is the restriction factor which is used to control the speed and the weight. The c_1 and c_2 are learning factors, and $rand_1$ and $rand_2$ are random values between 0 and 1. The classification process of PSO-SVM as shown in figure 2:

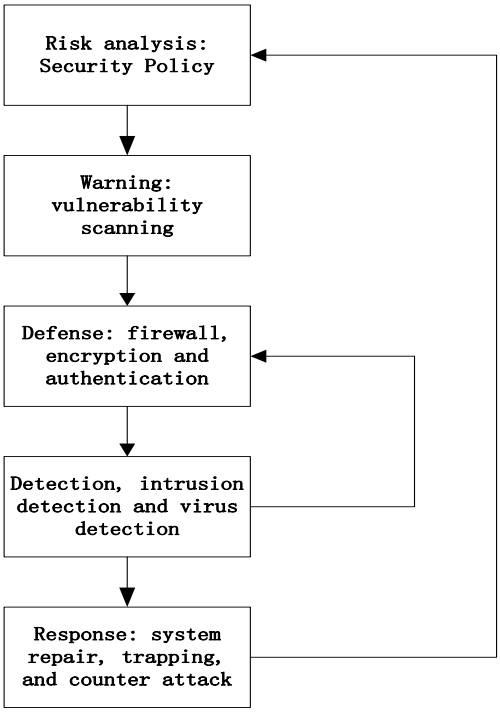

Figure 1: The network security defence system

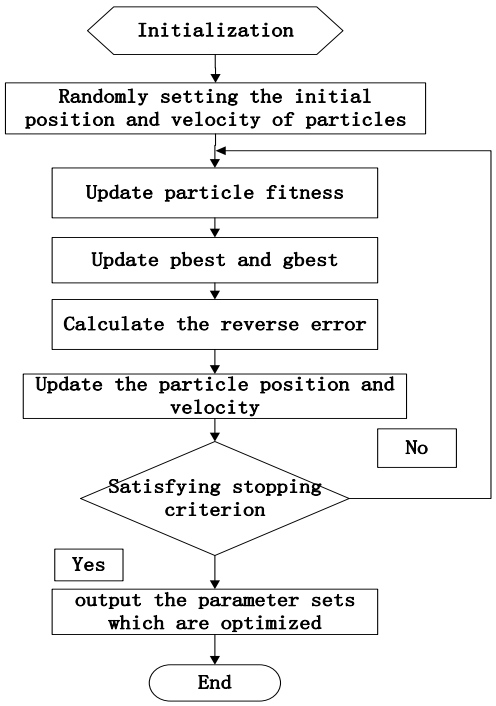

Figure 2: The classification process of PSO-SVM

4. Simulation experiment and result analysis

4.1 The experimental data

The training and testing datasets are stored by using text format in the KDD CUP99. At the same time, KDD CUP99 includes 4 main types of network attack types of MITLL intrusion detection data set, such as Denial-Of-Service (DoS), Surveillance or probe (Probe), User to Root (U2R), Remote to Local (R2L). In this data set, each row represents a record that contains 41 characteristic values. The 41 attributes are divided into four categories, which are the basic attribute set, the content attribute set, the traffic attribute set, and the host traffic attribute set. In this paper, the attack of Dos and Probe is our main test content, which are mainly related with the basic attribute set and the content attribute set. Therefore, we only introduce them which have 22 attributes, as shown in table 1 and table 2.

Table 1: The basic attribute set

Indicators	Description	type
1. duration	Connection time	continuous
2. protocol_type	TCP, UDP	discrete
3. service	HTTP, Telnet	discrete
4. flag	Connection status	discrete
5. src_bytes	The bytes from the source station to the destination station	continuous
6. dst_bytes	The bytes from the destination station to the source station	continuous
7. land	Whether it is connected to the same host	discrete
8. wrong_fragment	The number of wrong fragment	continuous
9. urgent	The number of urgent	continuous

Table 2: The content attribute set

Indicators	Description	type
10. hot	The number of "hot" indicator	continuous
11. num_failed_logins	The number of login failed	continuous
12. logged_in	The successful login is 1, otherwise is 0	discrete
13. num_compromised	The number of attack conditions satisfied	continuous
14. root_shell	For the super user is 1, otherwise is 0	discrete
15. su_attempted	Attempt to execute the "su root" is 1, otherwise is 0	discrete
16. num_root root	The number of access	continuous
17. num_file_creations	The number of file creation	continuous
18. num_shells	The number of Shell DOS prompt	continuous
19. num_access_files	The number of access control files	continuous
20. num_outbound_cmds	The number of FTP sessions with outbound command	continuous
21. is_hot_login	Login belongs to the "hot" list is 1, otherwise is 0	discrete
22. is _guest_login	Login belongs to the "guest" list is 1, otherwise is 0	discrete

4.2 The experiment steps and the result analysis

Step 1: According to the training data sets of DoS and Probe which are accounted for a larger proportion of all attacks, we use the IPSO-SVM classification algorithm for learning and generalization, and use the improved RBF kernel as our kernel function.

Step 2: Use the test data set of DoS and Probe to do the self-test and cross test.

Step 3: Set the statistics of performance evaluation as follow:

$$\text{total detection accuracy(TDA)} = \frac{\text{The number of correct classification}}{\text{The total number of samples}} \tag{8}$$

$$\text{false positive rate(FPR)} = \frac{\text{The number of samples to be false positive}}{\text{The number of normal samples}} \quad (9)$$

$$\text{detection rate (DR)} = \frac{\text{The number of abnormal samples are detected}}{\text{The number of abnormal samples}} \quad (10)$$

$$\text{average detection time(ADT)} = \frac{\text{The total detection time}}{\text{The total number of samples}} \quad (11)$$

Step 4: In the process of the experiment, we use the PSO method to determine the important parameters of the SVM algorithm, such as the RBF kernel parameter, penalty parameter and insensitive loss error. Then, the improved PSO method can find the optimal value of the SVM. At this time, the error sum of squares of the SVM model has a minimum value, and the model has a fast convergence speed.
Step 5: In order to reduce the randomness, we have 5 tests of intrusion detection, and compared with the traditional SVM classification method and the GA-SVM classification methods.
The test results of DR of Dos attacks as shown in Figure 3, and the test results of FPR of Probe attacks as shown in Figure 4.

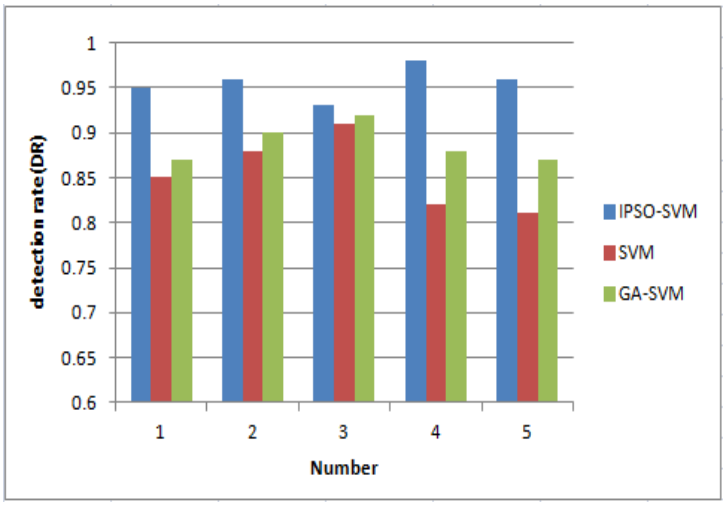

Figure 3: The comparison of detection rates for DoS attack

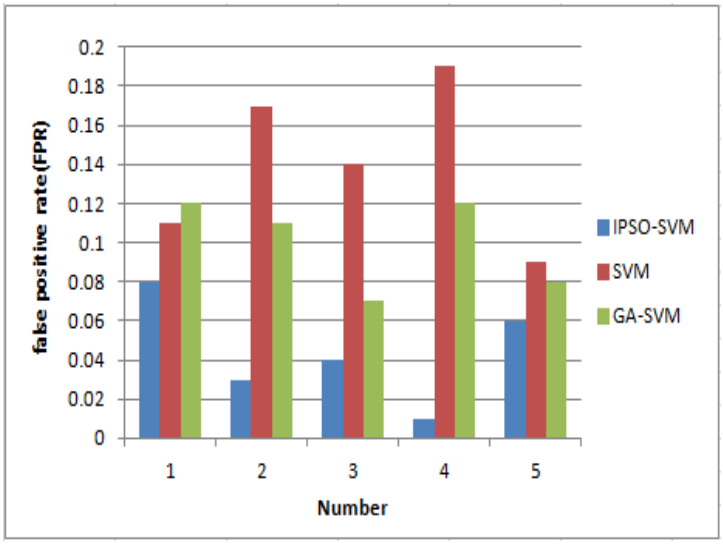

Figure 4: The comparison of false detection rates For Probe attack

From Figure 3 we can see, the DR of IPSO-SVM classification algorithm is significantly higher than that of the traditional SVM classification algorithm and the GA-SVM classification algorithm. The reason is that the momentum factor and adaptive rate accelerate the convergence of the PSO algorithm, so as to avoid falling into the local minimum. We can see from Figure 4 that FPR of IPSO-SVM classification algorithm is significantly lower than that of the traditional SVM classification algorithm and the GA-SVM classification algorithm. In summary, the overall performance of the proposed detection algorithm is very high. It has a strong ability to identify the characteristics of intrusion, and can provide the intrusion detection services for virtual environment.

5. Conclusions

In this article, we propose an intrusion detection method based on improved PSO-SVM. The support vector machine can ensure that classifier has high classification accuracies. Secondly, we use PSO method to determine the important parameters of the SVM algorithm, such as the RBF kernel parameter, penalty parameter and insensitive loss error. Then, the improved PSO method can find the optimal value of the SVM. At this time, the error sum of squares of the SVM model has a minimum value, and the model has a fast convergence speed. Finally, because the training data sets of DoS and Probe are accounted for a larger proportion of all attacks, we use the IPSO-SVM classification algorithm for them, and have a test to the intrusion detection. The experimental results show that the overall performance of the proposed detection algorithm is very high. It has a strong ability to identify the characteristics of intrusion, and can provide the intrusion detection services for virtual environment.

References

Denning D.E., 1987, An Intrusion-Detection Model [J]. IEEE Trans. Software Eng, 2-7.

Koc L., Mazzuchi T.A., Sarkani S., 2012, A network intrusion detection system based on a Hidden Naive Bayes multiclass classifier [J]. Expert Systems with Applications, 39(18): 13492-13500.

Lee W., Stolfo S.J., Mok K.W., 1999, A data mining framework for building intrusion detection models [C]. In Proceedings of the 1999 IEEE Symposium on Security and Privacy. Oakland, USA: IEEE, 120-132.

Li L., Zhang G.Y., Nie J.Y., 2012, The Application of Genetic Algorithm to Intrusion Detection in MP2P Network [J].Lecture Notes in Computer Science, 31(3): 390-397.

Li Q.H., Zhang D.Y., Sun Z.H., et al., 2003, Intrusion detection based on fuzzy reasoning driven by neural network [J]. Computer engineering, 29(19): 133-135.

Liu C., 2014, Network Intrusion Detection Model based on Artificial Fish Swarm Algorithm Optimizing Neural Network [J]. Computer Security, 7: 2-5.

Roesch M., 1999, Snort-Lightweight Intrusion Detection For Networks[J]. In Proceedings of LISA 99: 13th Systems Administration Conference, 2-4.

Shon T., Kovah X., Moon J., 2006, Applying genetic algorithm for classifying anomalous TCP/IP packets [J]. Neurocomputing, 69(16-18):2429-2433.

Wang G., Hao J.X., Ma J., 2010, A new approach to intrusion detection using artificial neural networks and fuzzy clustering [J]. Expert Systems with Application, 27(9): 6226-6232.

Xu Y.H., Li G.S., 2012, Incremental SVM Intrusion Detection Algorithm Based on Distance Weighted Template Reduction and Attribute Information Entropic [J]. Computer Science, 39(12):76-86.

Yang J., Chen X., Wan J.X., 2010, On Intrusion Detection Model Based on Network Protocol Analysis and Decision Tree Mining [J]. Computer Applications and Software, 27(2): 19-55.

Zhang J., Zulkernine M., Haque A., 2008, Random-Forests-Based network intrusion detection systems [J]. IEEE Transactions on System, Man, and Cybernetics, 38(5): 649-659.

Zhang Y.J., 2012, Intrusion detection method research based on incremental SVM [D].Shanxi: Xi'an University of Science And Technology.